molecular informatics
models – molecules – systems

Supporting Information

Supporting Information

69451 Weinheim, Germany

De novo design of bioactive small molecules by artificial intelligence

Daniel Merk, Lukas Friedrich, Francesca Grisoni, Gisbert Schneider*

Table of Contents

Experimental Procedures

Computational methods

Data preparation. Salts and stereochemistry information were removed, compound structures were represented in their neutral state. Molecular structures were represented as Simplified Molecular Input Line Entry System (SMILES) strings and converted to canonical SMILES with RDKit (Open-source cheminformatics; http://www.rdkit.org).

Generative machine learning model. All scripts were written in Python (Version 3.6), using RDKit (www.rdkit.org), Tensorflow (v1.2, www.tensorflow.org) and Keras (v2.0, https://keras.io) packages. The generative long short-term memory deep learning model trained on bioactive molecules from the ChEMBL database (ChEMBL22, pAffinity >6) was used as previously published.[1] The model was retrained (fine-tuning step) with a dataset containing 25 RXR and PPAR modulators (agonists and partial agonists) from literature and the authors' in-house collection, covering an activity range from 20 nM to 20 μM. For this fine-tuning step, the model was trained for five epochs with a softmax-temperature of 0.75 (see ref. [1] for details). Then, 1000 SMILES strings were sampled by fragment growing, choosing carboxylic acid (-COOH) as the start fragment. Growing was only allowed from the carbon atom of the start fragment. The retrieved 1000 SMILES strings were checked for validity (929 valid) and duplicates (900 uniques).

Chemical Space Analysis by Multi-Dimensional Scaling (MDS). Morgan fingerprints with an atom-centered radius of 0 to 4 bonds and a length of 1024 bit were computed using RDKit. Canonical MDS was computed using Jaccard-Tanimoto index as similarity metric. MDS was performed using the *cmds* module of MATLAB (v2017b, Mathworks, Natick, USA).

Similarity searching with holistic molecular descriptors. The similarity between the unique and valid molecules generated by the machine-learning model and the sets of RXR and PPAR actives was calculated using *Weighted Holistic Atom Localization and Entity Shape* (WHALES) descriptors (publication in preparation; see ref. 2 for a general description of the concept). Molecular geometry was optimized using the MMFF94[3] force field with 1000 iterations and 10 starting conformers for each compound; the minimum energy conformation was chosen for descriptor calculation. WHALES 3D descriptors were computed with in-house software, using Gasteiger-Marsili[4] partial charges as weighting scheme. RXR and PPAR query structures were retrieved from ChEMBL: (1) RXR binding: the 12 most potent ligands from ChEMBL; (2) RXR agonism: the top-4 agonists according to their EC_{50} value; (3) PPAR agonism: the top-4 agonists annotated in ChEMBL according to EC_{50} for each of the PPAR subtypes. For each dataset, every compound was used as query to perform similarity ranking on the basis of their Euclidean distance on Gaussian-normalized WHALES values. The results of the individual virtual screenings on each compound were merged according to the sum of their reciprocal ranks.[5]

Self-organizing map consensus for target prediction (SPiDER software): The bioactivity of all unique and valid molecules generated by the machine-learning model was predicted with the SPiDER software.[6] CATS2 descriptors[7] and the set of two-dimensional MOE descriptors (The Chemical Computing Group, Montreal, Canada; MOE2016.08; MOE descriptors KNIME node; forcefield: MMFF94*) were calculated for all generated molecules. The SPiDER results were filtered for compounds predicted to be active on RXR and/or a PPAR subtype, and compounds were ranked according to their *p*-value. Rank and *p*-value for each target of the selected designs are given in Table S1. The final list of high-scoring *de novo* designs is provided in Table S2.

Chemistry

General

All chemicals and solvents were reagent grade and used without further purification, unless specified otherwise. All reactions were conducted in oven-dried glassware under argon-atmosphere and in absolute solvents. NMR spectra were recorded on a Bruker AV 400 spectrometer (Bruker Corporation, Billerica, MA, USA). Chemical shifts (δ) are reported in ppm relative to TMS as reference; approximate coupling constants (J) are shown in Hertz (Hz). Mass spectra were obtained on an Advion expression CMS (Advion, Ithaka, NY, USA) equipped with an Advion plate express TLC extractor (Advion) using electrospray ionization (ESI). High-resolution mass spectra were recorded on a Bruker maXis ESI-Qq-TOF-MS instrument (Bruker). Compound purity was analyzed on a Varian ProStar HPLC (SpectraLab Scientific Inc., Markham, ON, Canada) equipped with a MultoHigh100 Phenyl 5 µ 240+4 mm column (CS-Chromatographie Service GmbH, Langerwehe, Germany) using a gradient (H_2O/MeOH 80:20+0.1% formic acid isocratic for 5 min to MeOH+0.1% formic acid after additional 45 min and MeOH+0.1% formic acid for additional 10 min) at a flow rate of 1 ml/min and UV-detection at 245 nm and 280 nm. All final compounds for biological evaluation had a purity > 95% (area-under-the-curve for UV_{245} and UV_{280} peaks).

*Synthesis and analytical characterization of **1**-**5** and precursors*

Methyl 3-(5-bromothiophene-2-carboxamido)benzoate (**8**)
2-Bromothiophene-5-carboxylic acid (**6**, 207 mg, 1.00 mmol, 1.00 eq), methyl 3-aminobenzoate (**7**, 166 mg, 1.10 mmol, 1.10 eq) and 4-DMAP (25 mg, 0.20 mmol, 0.20 eq) were dissolved in chloroform (10 ml, abs.), and EDC (202 mg, 1.30 mmol, 1.30 eq) was added. The resulting mixture was stirred 30 min at room temperature and then refluxed for 4 h. The reaction mixture was diluted with 20 ml ethyl acetate, poured to 20 ml 10% aqueous hydrochloric acid and phases were separated. The organic layer was washed with 20 ml 1 M aqueous sodium hydroxide solution, dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was washed with hexane/chloroform (9:1) to yield the title compound as colorless solid (220 mg, 65%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 3.86 (s, 3H), 6.80 (s, 1H), 7.04 (d, *J* = 4.0 Hz, 1H), 7.37 (t, *J* = 7.9 Hz, 1H), 7.55 (d, *J* = 4.0 Hz, 1H), 7.67 – 7.71 (m, 1H), 7.74 (dt, *J* = 1.3, 7.7 Hz, 1H), 7.85 (t, *J* = 2.0 Hz, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 52.32, 119.33, 121.09, 124.72, 125.83, 128.55, 129.35, 130.89, 131.04, 137.55, 140.35, 158.95, 166.59 ppm. MS (ESI-): m/z 337.7, 339.8.

Methyl 3-(5-phenylthiophene-2-carboxamido)benzoate (**10**)
8 (170 mg, 0.50 mmol, 1.00 eq) and benzeneboronic acid (**9**, 79 mg, 0.65 mmol, 1.30 eq) were dissolved in a mixture of dioxane (abs., 9 ml) and DMF (abs., 3 ml), caesium carbonate (488 mg, 1.50 mmol, 3.00 eq) was added and the mixture was stirred at room temperature for 30 min. Tetrakis(triphenylphosphin)palladium (29 mg, 0.025 mmol, 0.05 eq) was added and the mixture was stirred under reflux for 6 h. After cooling to room temperature, 25 ml 10% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate (9:1 to 4:1) as mobile phase and recrystallized from hexane/methylene chloride to yield the title compound as colorless solid (126 mg, 75%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 3.85 (s, 3H), 7.24 (d, J = 3.9 Hz, 1H), 7.27 – 7.41 (m, 4H), 7.54 – 7.59 (m, 3H), 7.75 (dt, *J* = 1.3, 7.8 Hz, 1H), 7.79 (s, 1H), 7.96 (ddd, *J* = 1.1, 2.3, 8.1 Hz, 1H), 8.06 (t, *J* = 1.9 Hz, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 52.29, 120.97, 123.65, 124.64, 125.59, 126.19, 128.83, 129.15, 129.33, 129.65, 131.01, 133.25, 137.23, 137.89, 150.21, 159.93, 166.65 ppm. MS (ESI+): m/z 339.0.

3-(5-Phenylthiophene-2-carboxamido)benzoic acid (**1**)
10 (85 mg, 0.25 mmol, 1.00 eq) was dissolved in a mixture of MeOH (1.0 ml), THF (1.0 ml) and water (2.0 ml) and potassium hydroxide (168 mg, 3.00 mmol, 12.0 eq) was added. The mixture was stirred at 70 °C under microwave irradiation for 30 min. After cooling to room temperature, 20 ml 5% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was recrystallized from hexane/ethyl acetate to yield the title compound as colorless solid (69 mg, 85%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 7.40 (d, *J* = 7.2 Hz, 1H), 7.48 (q, *J* = 7.9 Hz, 3H), 7.63 (d, *J* = 4.0 Hz, 1H), 7.68 (d, *J* = 7.8 Hz, 1H), 7.75 (d, *J* = 7.5 Hz, 2H), 8.02 (dd, *J* = 2.0, 8.1 Hz, 1H), 8.05 (d, *J* = 4.0 Hz, 1H), 8.34 (d, *J* = 2.1 Hz, 1H), 10.44 (s, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-d) δ = 126.18, 126.24, 129.54, 129.75, 130.98, 133.99, 134.25, 134.54, 135.74, 136.52, 138.11, 143.67, 144.15, 153.92, 165.09, 172.39 ppm. MS (ESI-): m/z 321.9. HRMS (ESI-): m/z 322.0543 calculated for $C_{18}H_{12}NO_3S$, found 322.0541 $[M-H]^-$.

5'-Fluoro-2'-hydroxy-2-(trifluoromethyl)-[1,1'-biphenyl]-4-carboxylic acid (**13**)
4-Bromo-3-trifluoromethylbenzoic acid (**11**, 270 mg, 1.00 mmol, 1.00 eq), 5-fluoro-2-hydroxybenzeneboronic acid (**12**, 155 mg, 1.00 mmol, 1.00 eq) and Cs_2CO_3 (980 mg, 3.00 mmol, 3.00 eq) were suspended in a mixture of toluene (abs., 8.0 ml) and ethanol (abs., 4.0 ml) and the mixture was stirred at room temperature for 30 min. Tetrakis(triphenylphosphine)palladium (58 mg, 0.05 mmol, 0.05 eq) was added and the mixture was stirred under reflux for 16 h. After cooling to room temperature, 25 ml 5% aqueous hydrochloric acid were added, phases were separated, and the aqueous layer was extracted three times with 30 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was

purified by column chromatography using hexane/ethyl acetate/acetic acid (79:19:2) and washed once with a cold mixture of CH_2Cl_2/hexane (1:4) to yield the title compound as colorless solid (96 mg, 32%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 6.75 – 6.83 (m, 2H), 6.93 (ddt, *J* = 3.4, 6.8, 8.9 Hz, 1H), 7.37 – 7.62 (m, 2H), 8.21 (dd, *J* = 1.7, 7.9 Hz, 1H), 8.42 (d, *J* = 1.7 Hz, 1H). ^{13}C NMR (101 MHz, chloroform-*d*) δ = 116.69, 122.02, 126.08, 128.41, 129.01, 132.06, 133.01, 133.09, 140.71, 148.67, 155.09, 157.46, 159.65, 170.28 ppm. MS (ESI-): m/z 298.9.

4-Fluorobenzyl 5'-fluoro-2'-((4-fluorobenzyl)oxy)-2-(trifluoromethyl)-[1,1'-biphenyl]-4-carboxylate (**14**)
13 (42 mg, 0.14 mmol, 1.00 eq) and 4-fluorobenzylbromide (**15**, 28 mg, 0.45 mmol, 3.00 eq) were dissolved in DMF (abs., 2 ml), potassium carbonate (63 mg, 0.45 mmol, 3.00 eq) was added and the mixture was stirred at 100°C under microwave irradiation for 2 h. After cooling to room temperature, 25 ml 5% aqueous hydrochloric acid were added, and the mixture was extracted three times with 30 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate (9:1) as mobile phase to yield the title compound as colorless oil (63 mg, 87%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 4.77 – 4.91 (m, 2H), 5.31 (s, 2H), 6.80 – 6.90 (m, 4H), 6.94 – 7.06 (m, 5H), 7.24 – 7.32 (m, 1H), 7.36 – 7.42 (m, 2H), 8.11 – 8.17 (m, 1H), 8.35 (d, *J* = 1.7 Hz, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 66.57, 70.24, 113.48, 115.27, 115.48, 115.57, 115.69, 115.78, 115.91, 117.30, 117.54, 127.55, 128.61, 129.22, 129.30, 129.78, 130.54, 131.43, 132.13, 132.58, 141.17, 151.63, 157.64, 165.05 ppm. MS (ESI-): m/z 514.8.

5'-Fluoro-2'-((4-fluorobenzyl)oxy)-2-(trifluoromethyl)-[1,1'-biphenyl]-4-carboxylic acid (**2**)
14 (51 mg, 0.10 mmol, 1.00 eq) was dissolved in a mixture of MeOH (1.0 ml), THF (1.0 ml) and water (2.0 ml) and potassium hydroxide (67 mg, 1.20 mmol, 12.0 eq) was added. The mixture was stirred at 70 °C under microwave irradiation for 30 min. After cooling to room temperature, 20 ml 5% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate/acetic acid (79:19:2) as mobile phase and crystallized from acetone/water to yield the title compound as colorless solid (34 mg, 83%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 4.80 – 4.92 (m, 2H), 6.81 – 6.92 (m, 4H), 6.96 – 7.06 (m, 3H), 7.35 (d, *J* = 7.9 Hz, 1H), 8.20 (d, *J* = 8.0 Hz, 1H), 8.43 (s, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 29.70, 70.26, 113.58, 115.50, 115.98, 117.52, 128.12, 128.61, 129.16, 131.14, 132.32, 132.72, 141.78, 151.62, 155.27, 157.20, 157.66, 163.55, 169.56 ppm. MS (ESI-): m/z 406.8. HRMS (ESI-): m/z 407.0712 calculated for $C_{21}H_{12}F_5O_3$, found 407.0721 $[M-H]^-$.

Methyl 4-bromo-2-hydroxybenzoate (**17**)
4-Bromosalicylic acid (**16**, 651 mg, 3.00 mmol, 1.00 eq) was dissolved in methanol (abs., 15 ml), sulfuric acid (>95%, 0.20 ml) was added and the mixture was stirred under reflux for 4 h. After cooling to room temperature 20 ml sat. sodium carbonate solution were added. The resulting cloudy aqueous solution was extracted twice with 50 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvent was evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate (9:1) as mobile phase to yield the title compound as colorless solid (610 mg, 88%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 3.88 (s, 3H), 6.95 (dd, *J* = 1.9, 8.5 Hz, 1H), 7.11 (d, *J* = 1.9 Hz, 1H), 7.61 (d, *J* = 8.5 Hz, 1H), 10.75 (s, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 52.50, 111.40, 120.87, 122.77, 129.97, 130.91, 162.01, 170.14 ppm. MS (ESI-): no molecular ion.

Methyl 4-bromo-2-(cyclopentyloxy)benzoate (**19**)
17 (230 mg, 1.00 mmol, 1.00 eq) and bromocyclopentane (**18**, 195 mg, 1.30 mmol, 1.30 eq) were dissolved in DMF (abs., 4.0 ml), sodium carbonate (212 mg, 2.00 mmol, 2.00 eq) was added and the mixture was stirred at room temperature until evolution of carbon dioxide had seized. The mixture was then stirred at 100°C under microwave irradiation for 2 h in a sealed vial. After cooling to room temperature, the reaction mixture was then poured to 25 ml 5% aqueous hydrochloric acid and extracted three times with 30 ml hexane at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane as mobile phase to yield the title compound as colorless oil (186 mg, 62%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 1.54 – 1.96 (m, 8H), 3.88 (s, 3H), 5.35 (td, *J* = 3.0, 6.1 Hz, 1H), 6.93 (dd, *J* = 1.9, 8.5 Hz, 1H), 7.09 (d, *J* = 1.9 Hz, 1H), 7.57 (d, *J* = 8.5 Hz, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 21.05, 23.75, 32.70, 78.86, 111.97, 120.77, 122.54, 129.62, 130.87, 162.08, 169.55 ppm. MS (ESI-): no molecular ion.

Methyl 3-(cyclopentyloxy)-3'-hydroxy-[1,1'-biphenyl]-4-carboxylate (**21**)
19 (150 mg, 0.50 mmol, 1.00 eq) and 3-hydroxybenzeneboronic acid (**20**, 69 mg, 0.50 mmol, 1.00 eq) were dissolved in a mixture of dioxane (abs., 6.0 ml) and DMF (abs., 3.0 ml), Cs_2CO_3 (490 mg, 1.50 mmol, 3.00 eq) was added and the mixture was stirred at room temperature for 30 min. Tetrakis(triphenylphosphine)palladium (29 mg, 0.025 mmol, 0.05 eq) was added and the mixture was stirred under reflux for 16 h. After cooling to room temperature, 25 ml 5% aqueous hydrochloric acid were added and the mixture was extracted three times with 30 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate/acetic acid (79:19:2) to yield the title compound as colorless oil (118 mg, 76%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 1.45 – 1.61 (m, 2H), 1.70 – 1.90 (m, 6H), 3.81 (s, 3H), 4.82 (tt, *J* = 3.3, 5.1 Hz, 1H), 6.79 (ddd, *J* = 1.0, 2.5, 8.1 Hz, 1H), 6.97 – 7.06 (m, 4H), 7.21 (t, *J* = 7.9 Hz, 1H), 7.74 (dd, *J* = 0.4, 7.9 Hz, 1H) ppm. ^{13}C NMR (101 MHz,

chloroform-*d*) δ = 20.77, 23.95, 32.82, 80.53, 113.45, 114.26, 115.16, 118.62, 119.29, 129.97, 132.15, 141.84, 146.13, 156.47, 158.05, 167.31, 177.83 ppm. MS (ESI-): m/z 310.7.

3-(cyclopentyloxy)-3'-hydroxy-[1,1'-biphenyl]-4-carboxylic acid (**3**)

21 (30 mg, 0.10 mmol, 1.00 eq) was dissolved in a mixture of THF (1.0 ml), methanol (1.0 ml) and water (2.0 ml) and potassium hydroxide (68 mg, 1.20 mmol, 12.0 eq) was added. The mixture was stirred at 70 °C under microwave irradiation for 15 min. After cooling to room temperature, 20 ml 5% aqueous hydrochloric acid were added and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate/acetic acid (79:19:2) as mobile phase and crystallized from methylene chloride/hexane to yield the title compound as colorless solid (21 mg, 70%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 1.63 – 1.87 (m, 4H), 1.91 – 2.07 (m, 4H), 5.10 (tt, *J* = 2.8, 5.8 Hz, 1H), 6.84 (ddd, *J* = 0.9, 2.5, 8.1 Hz, 1H), 7.01 (dd, *J* = 1.7, 2.5 Hz, 1H), 7.09 (ddd, *J* = 0.9, 1.7, 7.7 Hz, 1H), 7.13 (d, *J* = 1.6 Hz, 1H), 7.23 (dd, *J* = 1.6, 8.2 Hz, 1H), 7.28 (t, *J* = 7.9 Hz, 1H), 8.16 (d, *J* = 8.2 Hz, 1H), 10.78 (s, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 23.74, 32.97, 82.82, 112.40, 114.27, 115.62, 116.95, 119.79, 120.78, 130.30, 134.26, 141.28, 147.61, 156.14, 156.77, 171.89 ppm. MS (ESI-): m/z 296.8. HRMS (ESI-): m/z 297.1132 calculated for $C_{18}H_{17}O_4$, found 297.1134 [M-H]$^-$.

Methyl 3-(4-aminophenyl)propanoate (**23**)

3-(4-Aminophenyl)propionic acid (**22**, 0.50 g, 3.03 mmol) was dissolved in methanol (abs., 50 ml), sulfuric acid (>95%, 0.40 ml) was added and the mixture was stirred under reflux for 2 h. After cooling to room temperature, tert-butylmethyl ether (20 ml) was added to precipitate the product as hydrogen sulfate salt. The salt was dissolved in water (5 ml) and 20 ml sat. sodium carbonate solution were added. The resulting cloudy aqueous solution was extracted twice with 50 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvent was evaporated in vacuum to yield the title compound as brown oil (417 mg, 77%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 2.50 (dd, *J* = 7.1, 8.5 Hz, 2H), 2.77 (t, *J* = 7.8 Hz, 2H), 3.59 (s, 3H), 6.52 – 6.58 (m, 2H), 6.88 – 6.95 (m, 2H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 30.18, 36.16, 51.55, 115.30, 129.09, 130.54, 140.56, 173.57 ppm. MS (ESI+): no molecular ion.

Methyl 3-(4-(4'-chloro-[1,1'-biphenyl]-2-carboxamido)phenyl)propanoate (**25**)

Methyl 3-(4-aminophenyl)propanoate (**23**, 117 mg, 0.65 mmol, 1.30 eq), 2-(4-chlorophenyl)benzoic acid (**24**, 116 mg, 0.50 mmol, 1.00 eq), EDC (116 mg, 0.75 mmol, 1.50 eq) and 4-DMAP (12 mg, 0.10 mmol, 0.20 eq) were dissolved in $CHCl_3$ (abs., 10 ml) and stirred under reflux for 20 h. After cooling to room temperature, the reaction mixture was then poured to 50 ml 5% aqueous hydrochloric acid, phases were separated and the aqueous layer was extracted three times with 50 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography with hexane/ethyl acetate (4:1 to 2:1) as mobile phase to yield the title compound as colorless solid (138 mg, 70%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 2.54 (dd, *J* = 7.2, 8.3 Hz, 2H), 2.81 – 2.89 (m, 2H), 3.59 (s, 3H), 6.34 (d, *J* = 7.8 Hz, 1H), 6.78 – 6.82 (m, 1H), 7.02 – 7.06 (m, 1H), 7.07 – 7.11 (m, 2H), 7.11 – 7.15 (m, 2H), 7.20 – 7.26 (m, 2H), 7.29 – 7.37 (m, 2H), 7.38 – 7.44 (m, 2H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 30.34, 35.31, 35.77, 120.46, 127.19, 127.79, 128.04, 128.83, 129.75, 130.44, 134.44, 134.73, 135.57, 135.75, 136.57, 140.56, 151.55, 172.94, 173.28 ppm. MS (ESI+): m/z 394.1.

3-(4-(4'-chloro-[1,1'-biphenyl]-2-carboxamido)phenyl)propanoic acid (**4**)

25 (99 mg, 0.25 mmol, 1.00 eq) was dissolved in a mixture of THF (1.0 ml), methanol (1.0 ml) and water (2.0 ml) and potassium hydroxide (170 mg, 3.00 mmol, 12.0 eq) was added. The mixture was stirred at 70 °C under microwave irradiation for 30 min. After cooling to room temperature, 20 ml 5% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was recrystallized from hexane/acetic acid to yield the title compound as colorless solid (76 mg, 80%).
^{1}H NMR (400 MHz, chloroform-*d*) δ = 2.57 (t, *J* = 7.6 Hz, 2H), 2.84 (t, *J* = 7.6 Hz, 2H), 6.91 (s, 1H), 7.01 – 7.11 (m, 4H), 7.33 (t, *J* = 2.0 Hz, 4H), 7.39 (td, *J* = 1.4, 7.5 Hz, 1H), 7.46 (td, *J* = 1.5, 7.5 Hz, 1H), 7.70 (dd, *J* = 1.5, 7.6 Hz, 1H) ppm. ^{13}C NMR (101 MHz, chloroform-*d*) δ = 30.05, 35.31, 120.23, 128.13, 128.84, 129.07, 129.15, 130.03, 130.28, 130.71, 134.24, 135.57, 135.76, 136.63, 138.32, 167.21, 177.29 ppm. MS (ESI-): m/z 378.0. HRMS (ESI-): m/z 378.0902 calculated for $C_{22}H_{17}ClNO_3$, found 378.0909 [M-H]$^-$.

3',5'-dihydroxy-[1,1'-biphenyl]-4-carbaldehyde (**28**)

4-Formylbenzeneboronic acid (**26**, 150 mg, 1.00 mmol, 1.00 eq), 5-bromoresorcinol (**27**, 190 mg, 1.00 mmol, 1.00 eq) and Cs_2CO_3 (980 mg, 3.00 mmol, 3.00 eq) were suspended in a mixture of dioxane (abs., 5.0 ml) and DMF (abs., 1.0 ml) and the mixture was stirred at room temperature for 30 min. Tetrakis(triphenylphosphine)palladium (58 mg, 0.05 mmol, 0.05 eq) was added and the mixture was stirred under reflux for 4 h. After cooling to room temperature, 25 ml 10% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate (1:1) as mobile phase and crystallized from cold chloroform to yield the title compound as pale orange solid (131 mg, 61%).
^{1}H NMR (400 MHz, methanol-d_4) δ = 6.34 (t, *J* = 2.2 Hz, 1H), 6.62 (d, *J* = 2.2 Hz, 2H), 7.73 – 7.79 (m, 2H), 7.94 – 7.99 (m, 2H), 10.02 (s, 1H) ppm. ^{13}C NMR (101 MHz, methanol-d_4) δ = 102.33, 105.41, 120.01, 127.13, 129.74, 135.42, 141.59, 147.40, 158.80, 192.40 ppm. MS (ESI-): m/z 212.9.

3-(3',5'-dihydroxy-[1,1'-biphenyl]-4-yl)acrylic acid (**5**)

28 (42 mg, 0.20 mmol, 1.00 eq) and malonic acid (21 mg, 0.20 mmol, 1.00 eq) were dissolved in a mixture of pyridine (abs., 1.0 ml) and piperidine (0.10 ml). The mixture was stirred at 100°C under microwave irradiation for 30 min. After cooling to room temperature, 25 ml 10% aqueous hydrochloric acid were added, and the mixture was extracted three times with 25 ml ethyl acetate at a time. The combined organic layers were dried over magnesium sulfate and the solvents were evaporated in vacuum. The crude product was purified by column chromatography using hexane/ethyl acetate/acetic acid (49:49:2) as mobile phase and recrystallized from methylene chloride/hexane to yield the title compound as pale yellow solid (43 mg, 84%).

^{1}H NMR (400 MHz, methanol-d_4) δ = 6.18 (t, *J* = 2.2 Hz, 1H), 6.41 (d, *J* = 15.9 Hz, 1H), 6.47 (d, *J* = 2.2 Hz, 2H), 7.50 (d, *J* = 8.5 Hz, 2H), 7.54 (d, *J* = 8.5 Hz, 2H), 7.60 (d, *J* = 16.0 Hz, 1H) ppm. ^{13}C NMR (101 MHz, methanol-d_4) δ = 101.73, 105.07, 117.84, 126.93, 128.17, 133.44, 142.07, 143.14, 144.37, 158.67 ppm. MS (ESI-): m/z 254.8. HRMS (ESI-): m/z 255.0663 calculated for $C_{15}H_{11}O_4$, found 255.0664 $[M-H]^-$.

In vitro biological evaluation

Hybrid reporter gene assays for PPARα, PPARγ, PPARδ, RXRα, RXRβ and RXRγ

Plasmids: The Gal4-fusion receptor plasmids pFA-CMV-hPPARα-LBD, pFA-CMV-hPPARγ-LBD, pFA-CMV-hPPARδ-LBD, pFA-CMV-hRXRα-LBD, pFA-CMV-hRXRβ-LBD and pFA-CMV-hRXRγ-LBD coding for the hinge region and ligand binding domain (LBD) of the canonical isoform of the respective nuclear receptor have been reported previously. pFR-Luc (Stratagene) was used as reporter plasmid and pRL-SV40 (Promega) for normalization of transfection efficiency and cell growth.

Assay procedure: HEK293T cells were grown in DMEM high glucose, supplemented with 10% FCS, sodium pyruvate (1 mM), penicillin (100 U/ml) and streptomycin (100 µg/ml) at 37 °C and 5% CO_2. The day before transfection, HEK293T cells were seeded in 96 well plates ($2.5 \cdot 10^4$ cells/well). Before transfection, medium was changed to OptiMEM without supplements. Transient transfection was carried out using Lipofectamine LTX reagent (Invitrogen) according to the manufacturer's protocol with pFR-Luc (Stratagene), pRL-SV40 (Promega) and pFA-CMV-NR-LBD. 5 h after transfection, medium was changed to OptiMEM supplemented with penicillin (100 U/ml), streptomycin (100 µg/ml), now additionally containing 0.1% DMSO and the respective test compound or 0.1% DMSO alone as untreated control. Each concentration was tested in duplicates and each experiment was repeated independently at least two times. Following overnight (12-14 h) incubation with the test compounds, cells were assayed for luciferase activity using Dual-Glo™ Luciferase Assay System (Promega) according to the manufacturer's protocol. Luminescence was measured with an Infinite M200 luminometer (Tecan Deutschland GmbH). Normalization of transfection efficiency and cell growth was done by division of firefly luciferase data by *Renilla* luciferase data and multiplying the value by 1000 resulting in relative light units (RLU). Fold activation was obtained by dividing the mean RLU of a test compound at a respective concentration by the mean RLU of untreated control. All hybrid assays were validated with known reference agonists (PPARα: GW7647; PPARγ: pioglitazone; PPARδ: L165,041; RXRα/β/γ: bexarotene) which yielded EC_{50} values in agreement with literature. The assay was repeated in absence of a hybrid receptor by only transfecting pFR-Luc and pRL-SV40 for each active compound at a concentration around EC_{80} on a PPAR or RXR to exclude unspecific effects.

Supporting Figures and Tables

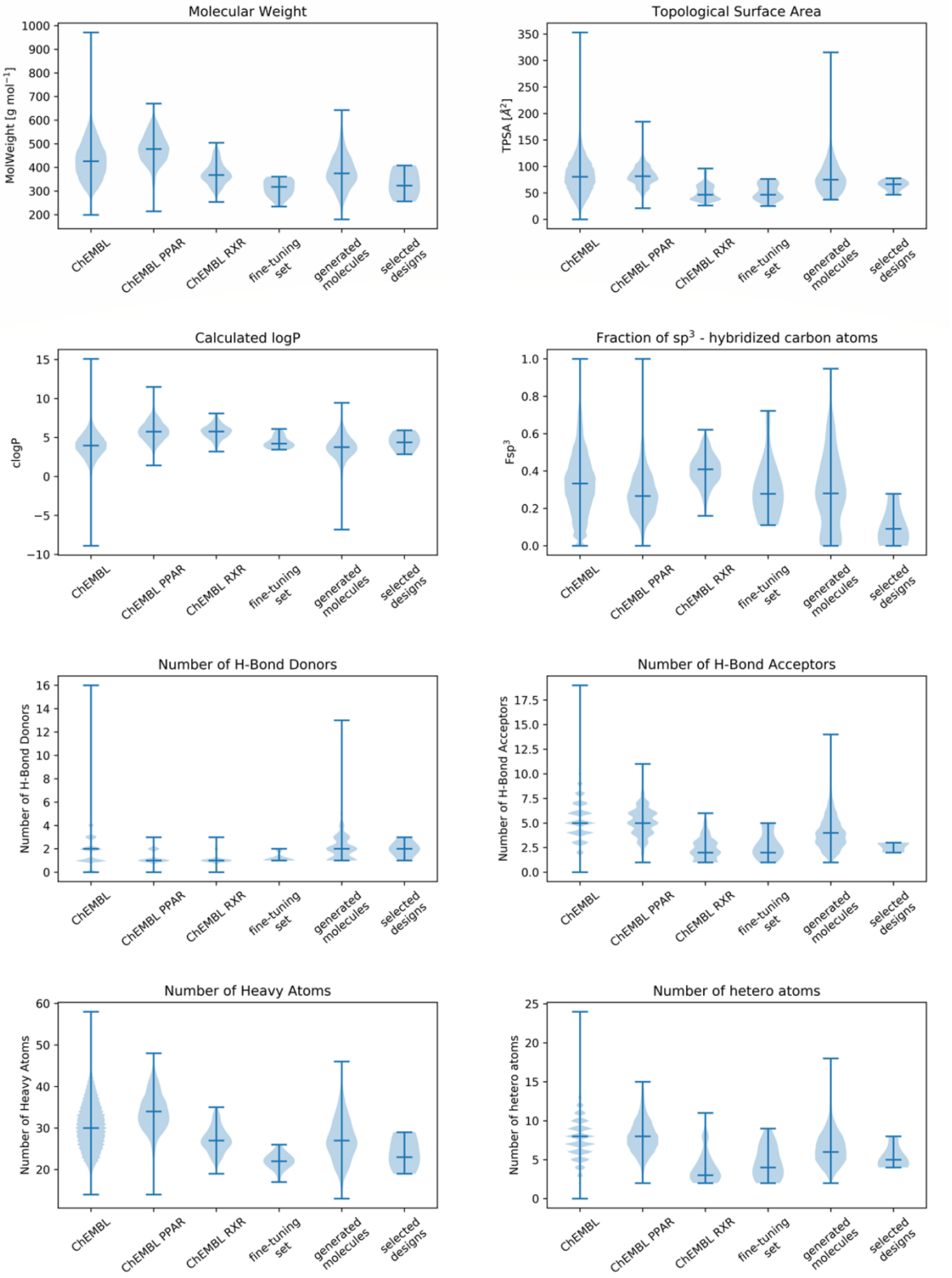

Figure S1. Characteristics of the training data (ChEMBL), ChEMBL RXR and PPAR actives, the fine-tuning set (25 actives), the generated molecules (900 uniques) and the synthesized designs **1**-**5**. Properties were calculated with RDkit. Violin plots show the median and extreme values.

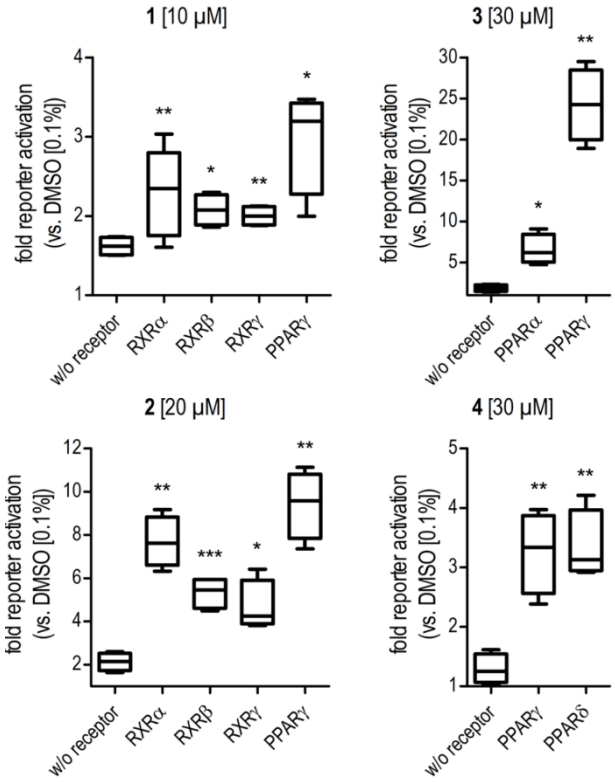

Figure S2. Validation of RXR and PPAR agonistic activity of designs **1**-**4**: **1**-**4** trans-activated reporter gene (firefly luciferase) expression via Gal4-hybrid receptors containing an RXR or PPAR ligand binding domain whereas there was no reporter transactivation in absence of a hybrid receptor (values are mean ± SEM fold activation compared to 0.1% DMSO; n≥3; t-test versus transactivation in absence of hybrid receptor: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Table S1. Ranks of the designs **1**-**5** for potential RXR and PPAR agonism by computational methods. SPiDER: *consensus* between a self-organizing map (SOM) on 2D-pharmacophoric features (CATS2) and a SOM on molecular properties (MOE). Similarity search: similarity-based virtual screening using sets of RXR and PPAR actives derived from ChEMBL using 3D shape- and charge-based descriptors (WHALES). Compound **1** was generated and ranked as 2,6-difluorophenyl derivative; for synthetic feasibility, the phenyl derivative **1** was prepared and tested.

	1*	**2**	**3**	**4**	**5**
SPiDER - RXR (*p*-value)	>100	>100	2 (0.052)	19 (0.125)	>100
SPiDER - PPAR (*p*-value)	>100	64 (0.012)	>100	43 (0.009)	19 (0.004)
WHALES - RXR binders	48	13	>100	>100	3
WHALES - RXR agonists	>100	>100	27	37	>100
WHALES - PPAR	48	74	>100	>100	>100

Table S2: Top-49 compounds contained in at least two Top-100 lists of in silico screenings for RXR and PPAR agonism.

O=C([O-])c1ccc(-c2noc(-c3ccc(F)cc3)n2)cc1
O=C(NCCOc1ccc(Cl)cc1)NC1CCC(OC(=O)[O-])CC1c1ccccc1
O=C([O-])c1ccc2c(c1)c1ccc(OCCCCc3ccccc3)cc1c1cnn(Cc3ccccc3)c21
O=C([O-])c1cccc(C(CC[NH+]2CC[NH+](CCOC(c3ccccc3)c3ccccc3)CC2)C(Cc2ccccc2)C(F)(F)F)c1
O=C([O-])c1ccc(NC(=O)c2cc3ccccc3[nH]2)cc1
O=C([O-])c1ccc(NC(=O)c2cccc(C(F)(F)F)c2)cc1
O=C([O-])c1cccc(Nc2nccc(-c3ccccc3)n2)c1
O=C([O-])c1ccc(-c2cccc(O)c2)cc1OC1CCCC1
O=C([O-])C1CC(OC(c2ccc(F)cc2)c2ccc(F)cc2)C[NH2+]1
O=C([O-])/C=C/c1ccc(Oc2ccc(-c3ccccc3)cc2)c(Cl)c1
O=C(O)c1ccc(-c2cccc(CCCc3ccccc3)n2)c(Cl)c1
O=C([O-])c1ccc(-c2cccc(CCCc3ccccc3)n2)c(Cl)c1
O=C(O)c1ccc(-c2ccc(CCCCc3ccccc3)cc2)nc1
O=C([O-])c1ccc(Nc2nccc(-n3nnc4ccccc43)n2)cc1
O=C([O-])c1ccc(-c2cc(C(F)(F)F)nn2-c2ccc(Cl)cc2Cl)cc1
O=C([O-])CCCNc1nc2ccccc2n1Cc1ccc(C(=O)[O-])cc1
O=C([O-])c1cc(Cl)c(Oc2ccc(OCc3ccccc3)cc2)c(Br)c1
O=C([O-])c1ccc(CNc2nccc(-c3cc4ccccc4s3)n2)cc1
O=C([O-])c1cccc(-c2ccc(NC(=O)c3cccc([N+](=O)[O-])c3)cc2)c1
O=C(O)CCc1ccc(NC(=O)c2ccccc2-c2ccc(Cl)cc2)cc1
O=C([O-])c1ccc(C2=CC[NH+](CCCNC(=O)c3ccc(Cl)cc3)C2)cc1
O=C(O)c1cccc(-c2cc(-c3ccccc3)n(CC3CCCCC3)n2)c1
O=C([O-])c1ccc2c(-c3ccccc3)c(C(F)(F)F)c3ccccc3c2c1
O=C([O-])c1ccc(Nc2ncc(F)c(Nc3ccc4c(c3)OCO4)n2)cc1
O=C([O-])/C=C/c1cccc(C(O)C[NH2+]CCOc2ccc3ccccc3c2)c1
O=C(Nc1ccc(C(=O)[O-])cc1)C(=O)c1ccc(Oc2ccc(Cl)cc2)cc1
O=C([O-])c1ccc(-c2ccc(C[NH2+]CC(O)COc3ccccc3)cc2)cc1
O=C([O-])c1ccc(-c2cc3c(c(OCCC[NH+]4CCCCC4)c2)CCC3)cc1
O=C([O-])c1ccc(-c2cc(F)ccc2OCc2ccc(F)cc2)c(C(F)(F)F)c1
O=C([O-])c1ccc2[nH]c(-c3ccc(C(=O)NCCc4ccccc4)cc3)nc2c1
O=C([O-])CC1CCC(CNC(=O)c2ccc(-c3cc(C(=O)[O-])ccc3F)cc2)CC1
O=C(O)c1ccc(-c2ccccc2Oc2ccc(Oc3ccc(Cl)cc3)cc2)cc1
O=C([O-])c1ccc(-c2ccccc2Oc2ccc(Oc3ccc(Cl)cc3)cc2)cc1
O=C(O)c1ccc(Oc2ccc(C3(c4ccc(F)cc4F)CCOCC3)cc2)cc1
O=C(O)c1ccccc1OCCCc1ccc(OCc2ccccc2C(F)(F)F)cc1
O=C(O)c1ccc(N2CCN(C(=O)CCCCCCCCc3ccccc3)CC2)cc1
O=C(O)C1Cc2ccccc2C(=O)N1Cc1ccc(Oc2cccc3[nH]ccc23)cc1
O=C([O-])c1ccc(OCCc2n[nH]c3ccc(-c4ccc5ccccc5c4)cc23)cc1
O=C([O-])c1cn(Cc2ccccc2)c2ccc(OCc3cccc(OC(F)(F)F)c3)cc12
O=C([O-])CCCCCNc1nc(O)c2ncn(Cc3ccc(-c4ccccc4)cc3)c2n1
O=C([O-])c1ccc(C(=C2CC3CCC(C2)[NH+]3CCc2ccccc2)c2ccccc2)cc1
Nc1nccc(-c2c(-c3ccccc3)cc(C(=O)[O-])cc2-c2ccc3[nH]ncc3c2)c1Cl
O=C([O-])NCC1CN(c2ncc(-c3cccc(F)c3)c(-c3ccc(C(F)(F)F)cc3)n2)CCO1
O=C([O-])c1ccc(C(O)C[NH+]2CCC(c3cn(-c4ccc(F)cc4)c4ccccc34)CC2)cc1
O=C([O-])C1=C(c2ccc(F)cc2)c2ccc(O)cc2C2=CCCN(C(=O)c3ccccc3)CC21
O=C(O)c1cccc2c1-c1cccc3c1C(C2)N(C(=O)CCCCOCc1ccccc1)CC3
O=C([O-])c1ccc(-c2ccc3ncnc(Nc4ccc(OCc5ccccc5)c(Cl)c4)c3c2)cc1
O=C([O-])CCCCCCCCCCCCCCCC[NH+]1CCN(S(=O)(=O)c2ccc(OCCO)cc2)CC1
O=C([O-])c1c(-c2ccc(OC(F)(F)F)cc2)nn2c1N=C(c1ccc(Cl)cc1)C1OC(CO)C(O)C(O)C12

References

[1] A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700111.

[2] F. Grisoni, D. Reker, P. Schneider, L. Friedrich, V. Consonni, R. Todeschini, A. Koeberle, O. Werz, G. Schneider, *Mol. Inf.* **2017**, *36*, 1600091.

[3] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490–519.

[4] J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219–3228.

[5] B. Chen, C. Mueller , P. Willett, *Mol. Inf.* **2010**, *29*, 533–541.

[6] D. Reker, T. Rodrigues, P. Schneider, G. Schneider, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4067–4072.

[7] M. Reutlinger, C. P. Koch, D. Reker, N. Todoroff, P. Schneider, T. Rodrigues, G. Schneider, G. *Mol. Inf.* **2013**, *32*, 133–138.

Author Contributions

G.S. designed the study; D.M. ranked, selected, synthesized, characterized and tested the designs; L.F. fine-tuned the model, generated and analyzed the designs; F.G. ranked and analyzed the designs; D.M. and G.S. analyzed the in vitro data; D.M. and G.S. supervised the project; D.M. and G.S. wrote the manuscript; all authors approved the final version of the manuscript.